MBMG Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 9: 149–173 (2025)
DOI: 10.3897/mbmg.9.137772

Research Article

# An annotated reference library for supporting DNA metabarcoding analysis of aquatic macroinvertebrates in French freshwater environments

Paula Gauvin[1,2], David Eme[1,2], Isabelle Domaizon[1,2], Frédéric Rimet[1,2]

1 UMR CARRTEL, Université Savoie Mont-Blanc, INRAE, Pole R&D ECLA, 75bis av. de Corzent – CS 50511, FR-74200, Thonon-les-Bains, France
2 RiverLY Research Unit, National Research Institute for Agriculture Food and Environment, (INRAE), Villeurbanne, France
Corresponding author: Paula Gauvin (Paula.gauvin@inrae.fr)

## Abstract

Freshwater ecosystems are increasingly threatened by human activities, leading to biodiversity loss and ecosystem degradation. Effective biodiversity monitoring, particularly through the use of aquatic macroinvertebrates as bioindicators, is crucial for assessing ecological health. While traditional morphological methods face limitations, DNA metabarcoding offers higher accuracy and efficiency in species identification using environmental DNA. However, the success of metabarcoding is contingent on the quality of reference libraries, which are often incomplete or biased. This study aimed to construct and share a comprehensive COI-based DNA barcode library for freshwater macroinvertebrates in France, specifically targeting short gene regions amplified with fwhF2/fwhR2N primers, suitable for degraded DNA. A list of species occurring in French freshwater ecosystems was established from official national checklists and Alpine lake surveys. The resulting library was analysed for taxonomic completeness, barcode coverage and cryptic diversity. The checklist consisted of 2,841 species across 10 phyla, for which 56% had at least one COI-5P sequence available in the Barcode of Life Data System (BOLD). The analysis of cryptic diversity, based on Barcode Index Numbers (BINs) highlighted a potential high rate of cryptic diversity, although it might have been overestimated due to the wide geographic origin of the sequences. Alignment challenges with the primers were identified for certain taxa, particularly amongst Coleoptera, Diptera and Malacostraca. The genetic diversity approached by the number of haplotypes per species highlighted that most of the species have limited diversity, with only three species having more than 100 haplotypes. Finally, this study showed that a total of 57 haplotypes were shared amongst 116 distinct species. This work emphasises the need for expanded sequencing efforts to improve barcode coverage and highlighted the pitfalls associated with the use of these primers for further biodiversity assessment of macroinvertebrates with DNA.

**Key words:** COI, cryptic diversity, freshwater, fwhF2 primers, gap analysis, haplotype diversity, macroinvertebrates, metabarcoding



**Academic editor:** Alexander Weigand
**Received:** 24 September 2024
**Accepted:** 18 February 2025
**Published:** 7 April 2025

**Citation:** Gauvin P, Eme D, Domaizon I, Rimet F (2025) An annotated reference library for supporting DNA metabarcoding analysis of aquatic macroinvertebrates in French freshwater environments. Metabarcoding and Metagenomics 9: e137772. https://doi.org/10.3897/mbmg.9.137772

## Introduction

Freshwater ecosystems face escalating anthropogenic pressures threatening their functionality (Reid et al. 2019) and leading to significant biodiversity loss (Young et al. 2016; Borgwardt et al. 2019). In this context, biodiversity monitoring is crucial to provide relevant ecological diagnoses and support the management and conservation of these ecosystems and the vital services they provide. Aquatic macroinvertebrates are widely used as bioindicators due to their abundant presence, high species diversity and sensitivity to both anthropogenic and natural disturbances (Hering et al. 2006). These organisms primarily inhabit the littoral zone and play pivotal roles in community assembly and food web dynamics, contributing significantly to beta diversity and metacommunity structure across diverse aquatic habitats (May 2019). Consequently, they serve as key indicators in environmental monitoring programmes, facilitating assessments of the ecological status of aquatic ecosystems as mandated by the EU Water Framework Directive (WFD) and national legislation (Mondy et al. 2012) and are key indicators for assessing the effectiveness of restoration measures (e.g. Van Der Lee et al. (2024)). The standardised methodology for macroinvertebrate biomonitoring in Europe is often based on morphological identification through binocular microscopy. This methodology encounters significant challenges, including the labour-intensive and expensive nature of collecting and sorting individual benthic invertebrates, which hinders its broader adoption in routine biomonitoring (Bonada et al. 2006; Blackman et al. 2019). Moreover, routine morphological identification poses drawbacks, such as high expertise demands and low taxonomic resolution (Leese et al. 2016, 2018; Bean et al. 2017; Hering et al. 2018). Notably, many taxa can only be identified at limited taxonomic resolution (Caesar et al. 2006) without a timely effort of highly skilled taxonomists. These challenges underscore the need for innovative approaches to enhance the efficiency and accuracy of macroinvertebrate biomonitoring practices. The recent advent of DNA metabarcoding techniques, integrating amplicon barcoding with high-throughput sequencing, represents a promising advancement in biomonitoring applications (Deiner et al. 2017; Pawlowski et al. 2018; Carraro and Altermatt 2024), providing a valuable complement to morphology-based approaches for species identifications. Metabarcoding finds application in analysing environmental DNA (eDNA) samples obtained from water, biofilm or sediment (Valentini et al. 2016; Sakata et al. 2020; Rivera et al. 2021) or DNA extracts derived from a homogenate of the sample's fauna (Taberlet et al. 2012). These molecular innovations enable simultaneous processing of multiple samples, identification of small taxa, immature or larval stages and offer increased sensitivity and specificity, often revealing hidden diversity, while enhancing time and cost effectiveness (Shokralla et al. 2012; Pochon et al. 2015; Holman et al. 2019). These advantages facilitate direct comparison amongst sites and studies and enable higher spatial-temporal frequency in monitoring due to increased throughput (Bush et al. 2019). The efficiency of macroinvertebrates DNA metabarcoding relies heavily on the effectiveness of primer sets across a broad taxonomic spectrum. The recovery rate of taxa using metabarcoding is contingent upon various factors, including the taxonomic resolution of the gene marker employed (e.g. COI or ribosomal markers like 16S, for example, Elbrecht et al. (2016)), amplicon length (Meusnier et al. 2008), primer universality and the number of primer pairs used to amplify target taxonomic groups (Elbrecht and Leese 2015; Gibson

et al. 2015). Specifically, a segment of the cytochrome c oxidase subunit I (COI) gene has emerged as the standard DNA barcoding marker for most animal groups (Hebert et al. 2003a), with over 95% of species in diverse animal groups exhibiting distinctive COI sequences in test assemblages (Hebert et al. 2003b, 2004). Given the considerable phylogenetic diversity amongst macroinvertebrates, the high taxonomic resolution and existing reference data for the COI marker make it a judicious choice for metabarcoding freshwater macroinvertebrate communities (Ratnasingham and Hebert 2007; Andújar et al. 2018). Environmental DNA released by target taxa can degrade rapidly (Seymour et al. 2018). Consequently, for amplifying degraded DNA that is extracted from water samples for example, targeting a short COI marker region is suggested to enhance amplification success (Barnes et al. 2014).The effectiveness of DNA metabarcoding programmes relies on open, comprehensive and accurate reference sequence libraries (Briski et al. 2016; Oliveira et al. 2016; Weigand et al. 2019), ensuring precise taxonomic assignment (Richardson et al. 2018; Rodriguez-Ezpeleta et al. 2021). In that sense, incomplete DNA barcode libraries, as in poorly-represented species, pose a significant challenge, often leading to false negatives and compromising biodiversity assessments (Ardura 2019; Leite et al. 2020; Duarte et al. 2021). Therefore, evaluating these gaps and the quality of sequence data in reference libraries is imperative for the effective implementation of DNA-based tools in biodiversity assessments (Duarte et al. 2020). Given the significance of reference libraries completeness, primarily those that are pertinent in the context of the WFD, numerous studies have been conducted to assess their representativeness by comparing them with lists of described species (Trebitz et al. 2015; Weigand et al. 2019; Duarte et al. 2020; Leite et al. 2020; Specchia et al. 2020; Csabai et al. 2023). Despite efforts to assess their representativeness, biases in taxonomic coverage persist within reference libraries (Li et al. 2019; Weigand et al. 2019). Indeed, numerous studies have emphasised the construction of tailored reference libraries to suit the geographic context of the research area (Ficetola et al. 2021: Mugnai et al. 2023) and emphasised the importance of such databases to be refined with custom sequences specific to local study areas and free from unexpected taxa. Additionally, Abad et al. (2016) and Schenekar et al. (2020) highlighted the value of possessing DNA barcodes for local species. These studies have demonstrated that DNA barcodes from local organisms can improve the accuracy of taxonomic assignments and reveal previously unrecognised biodiversity, leading to adjustments in taxonomic classifications amongst species. Our study is in line with the recommendations proposed by Blackman et al. (2023), which advocate for compiling a comprehensive list of target species within the study area and assembling accurate sequences pertinent to those species. Our main objective is to construct an open-access DNA barcode library for metabarcoding studies focused on French freshwater macroinvertebrates. To achieve this, we first compiled a list of macroinvertebrate species known to be present in French freshwater ecosystems. Then, we aimed to identify the gaps in barcode coverage within this new reference library and assess cryptic diversity, based on Barcode Index Numbers (BINs) (Ratnasingham and Hebert 2013). Finally, we focused on a short region of the COI gene which is highly effective when targeting degraded DNA of freshwater macroinvertebrate fauna (fwhF2/fwhR2N, 205 bp, Vamos et al. (2017)) and assessed its ability to amplify barcodes of the built reference library. This should thereby provide insight into the limitations of these primers and facilitate metabarcoding data interpretation for future users.

## Material and methods

### Checklist constitution for French metropolitan freshwater macroinvertebrates species

In order to establish the most up-to-date checklist of macroinvertebrates species recognised and described occurring in freshwater aquatic habitats in metropolitan France, three checklists were complied. The first one comes from PERLA, an interactive tool managed by a French regional environmental agency (DREAL Auvergne-Rhône-Alpes) accessible at http://www.perla.developpement-durable.gouv.fr/index.php. PERLA serves as a national comprehensive checklist for water managers and encompasses larvae, nymphs and adults across various taxonomic groups, including insects, molluscs, crustaceans and worms, amongst others, found in rivers and aquatic ecosystems. In the manuscript, we will refer to this checklist as the 'French aquatic ecosystems checklist'. The second checklist is coming from macroinvertebrate surveys conducted in four Alpine lakes (Lake Geneva, Lake Annecy, Lake Bourget and Lake Tignes) from 2015 to 2022. In the manuscript, we will refer to this checklist as the 'French Alpine lakes checklist'. The third checklist we used was from a French NGO, Opie-benthos (https://www.opie-benthos.fr/opie/monde-des-insectes.html), studying freshwater insect's taxonomy and diversity. Currently, Opie-benthos gathers, curates and maintains certainly the most up-to date freshwater insect species list for the French metropolitan territory. In the manuscript, we will refer to this checklist as the 'French aquatic insect's species list'. These three checklists were merged into a single one. When the taxonomic resolution of taxa was limited to a level above species (genus, family, class, phylum), we consulted the taxonomic referential of the National Inventory of Natural Heritage (INPN - Inventaire national du patrimoine naturel) (TAXREF v.17.0 2024, available at https://inpn.mnhn.fr/telechargement/referentielEspece/taxref/17.0/menu) to select species from those ranks. The phylum Acanthocephala and Rotifera, the classes Copepoda and Ostracoda (Arthropoda, Crustacea) and the families Succineidae (Arthropoda, Insecta, Diptera) and Hydrachnidae (Arthropoda, Arachnida, Trombidiformes), initially listed in French aquatic ecosystems checklist, were omitted from our search from TAXREF v.17.0 2024 as they are not considered as freshwater macroinvertebrates according Tachet et al. (2000). Once those macroinvertebrates species were retrieved for each taxonomic level without species identification, several filters from TAXREF were applied to select only French metropolitan freshwater species from this taxonomic referential. First, we selected only the species occurring in the following habitats: freshwater, marine and freshwater, brackish water and continental (land and freshwater). Then, from this updated list, only species occurring in metropolitan France were kept. Finally, species characterised by one those eight statuses over 15 statuses in total were selected (present (native or undetermined), endemic, sub-endemic, cryptogenic, introduced, invasive introduced, non-established introduced (including cultivated/domestic) and occasional). The checklists described above and the summarised species list resulting from these compilations, inclusive of all species from the mentioned inventories, is referred to as the "French metropolitan freshwater macroinvertebrates' species list" and all together are available at https://doi.org/10.57745/LMOXEW in the 'Checklists composition data' file for future reference.

## Sequences origin and cleaning steps

All the sequences for those species listed in the French metropolitan freshwater macroinvertebrates' species list, whatever the gene, were downloaded from October 2023 to May 2024, from the Barcode of Life Data System v.4 -Bold-repository (https://boldsystems.org/index.php/databases). We opted to utilise BOLD as the reference library instead of GenBank due to concerns regarding the latter's unverified submission process, which frequently results in misannotated sequences (Kozlov et al. 2016; Locatelli et al. 2020; Steinegger and Salzberg 2020). The absence of sequences for a given species may indicate either that the search in BOLD returned "Unmatched terms" or that, while the species was found, no sequences were publicly available, rendering them private sequences (regardless of the reason for their absence, they were all defined as "not available sequences" for this analysis). Before concluding that a species from the French aquatic ecosystems' checklist and French Alpine lakes checklist lacked sequences in BOLD, synonyms were searched using INPN. The details of taxa resulting from this taxonomic harmonisation are available in Suppl. material 1: table S1. From this file containing all the sequences retrieved (whatever the gene) for the queried species of the French metropolitan freshwater macroinvertebrates' species list, the COI-5P sequences belonging to the COI gene were selected. All the remaining sequences were aligned with the fwhF2/fwhR2N primer pair using MAFFT version 7 (Auto) (https://mafft.cbrc.jp/alignment/server/), by genus or species group (i.e. all sequences assigned to the same genus or species were aligned together). Then, sequences characterised by gaps, insufficiencies in length relative to the COI barcode produced by the fwhF2/fwhR2N primer pair (shorter than 205 bp) or identical (same genetic sequence for one species) from the sequences file, were removed using Jalview (https://www.jalview.org/). Due to the possibility of identical sequences being shared amongst different species, certain species may have been excluded during this stage, as a result of sequence overlap with other taxa. To address this issue, the list of species obtained from this step was cross-referenced with the initial species list following the alignment process. This comparison identified any missing species, whose sequences were then re-aligned individually to ensure their accurate representation. This final reference library containing only genetic sequences capable of being aligned by the fwhF2/fwhR2N primer pair is named 'Aligned DNA library' hereafter. The reference barcode libraries (with all genes, COI-5P fragments only and the shorter COI-5P fragment matching fwhF2/fwhR2N primers) and a summary table of the number of sequences and Barcode Index Numbers (BINs) as potential proxy of cryptic species are available at https://doi.org/10.57745/LMOXEW for future reference.

## Graphical analysis

We firstly analysed the taxonomic composition of the French metropolitan freshwater macroinvertebrates' species list. Secondly, to highlight any gaps in the availability of sequences in our list of species, the taxonomic coverage of barcodes at various taxonomic levels was assessed. To estimate the importance of cryptic diversity for the different taxonomic groups, we used the number of Barcode Index Numbers (BINs) per species. We also reported the

number of BINs lumping sequences attributed to different species. To account for the unbalanced availability of the DNA sequences (and sampling effort) amongst species, we also evaluated the correlation between the number of BINs and sequences. Thirdly, we highlighted the taxonomic composition of species with COI-5P sequences that did not align with fwhF2/fwhR2N primers to reveal which species and taxonomic groups could be misrepresented in metabarcoding studies using that primer pair. Then, to explore the species haplotype diversity, the number of different haplotypes available relative to the number of species by phylum was examined. The categorisation of unique barcodes ranging from limited (< 5 haplotypes) to moderate (5–25 haplotypes) to good (> 25 haplotypes), was adopted from Trebitz et al. (2015) and the species with the highest number of haplotypes exceeding 100 were identified. Finally, in order to demonstrate the ability of the primers to discriminate each species by a unique barcode, groups of taxa sharing the same haplotype were identified.

## Results and discussion

### Taxonomic composition of the French metropolitan freshwater macroinvertebrates' species list, barcoding coverage and cryptic diversity

The French metropolitan freshwater macroinvertebrates' species list is composed of 10 phyla, 16 classes, 50 orders, 222 families, 670 genera and 2841 species (Table 1).

**Table 1.** Summary of taxonomic composition for each phylum of the French metropolitan freshwater macroinvertebrates' species list.

| Phylum | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|
| Annelida | 1 | 2 | 6 | 18 | 40 |
| Arthropoda | 4 | 20 | 135 | 508 | 2469 |
| Bryozoa | 2 | 2 | 6 | 7 | 11 |
| Cnidaria | 1 | 2 | 3 | 3 | 7 |
| Entoprocta | 1 | 1 | 1 | 1 | 1 |
| Mollusca | 2 | 11 | 24 | 52 | 195 |
| Nematoda | 2 | 9 | 42 | 68 | 99 |
| Nemertea | 1 | 1 | 1 | 1 | 1 |
| Platyhelminthes | 1 | 1 | 3 | 7 | 12 |
| Porifera | 1 | 1 | 1 | 5 | 6 |

Fig. 1 provides insights into the taxonomic composition, highlighting that the highest diversity is observed within the phylum Arthropoda, followed by Mollusca and Nematoda, with 2469, 195 and 99 species, respectively. On the other hand, taxa such as Entoprocta and Nemerta are represented by only one species (Fig. 1, Table 1).

Amongst the insects, Coleoptera dominate, followed by Diptera and Trichoptera, with 684, 630 and 527 species, respectively. The predominant phyla with the highest number of species in our checklist matches with those identified in a study by Specchia et al. (2020), who conducted a gap analysis of DNA barcodes

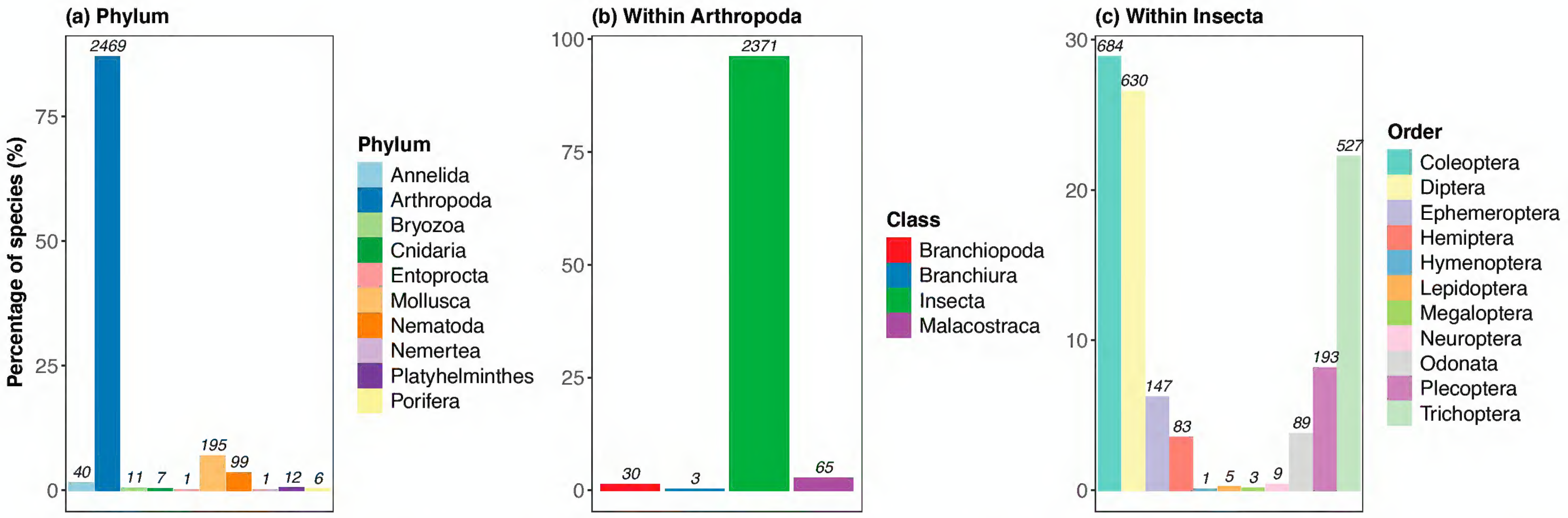


**Figure 1.** Taxonomic composition of the French metropolitan freshwater macroinvertebrates' species list. Percentage of the number of species according to phylum (**a**), class within the arthropod phylum (**b**) and order within the insect class (**c**). The number above each bar represents the total number of species affiliated to each taxonomic rank.

available in international repositories using the aquatic macroinvertebrate species checklist of a south-eastern Apulia region in Italy. Furthermore, our findings align with the prevailing understanding that Arthropoda and thus insects, represent the most diverse group of animals, exerting dominance in freshwater ecosystems inventories (Yeates and Wiegmann 1999; Grosberg et al. 2012; Dijkstra et al. 2014; Choudhary and Ahi 2015). Moreover, Diptera emerge as the most species-rich group utilised in biomonitoring across Europe, with chironomids recognised for their prevalence and diversity in freshwater habitats (Pinder 1986). Alongside Diptera, Coleoptera (beetles) stand out for their exceptionally high species numbers across various ecoregions and countries in Europe and serving as the most abundant group of aquatic insects (Jäch and Balke 2008; Short 2018). Following these orders, Trichoptera (caddisflies), Plecoptera (stoneflies) and Ephemeroptera (mayflies), collectively known as EPT, emerge as the next three orders in terms of species richness from our checklist. These organisms spend their immature stages in freshwater and are widely employed as biological indicators for freshwater quality assessment (Hering et al. 2004; Sweeney et al. 2011) and ecological investigations, demonstrating robust responses to pollution or climate change (Álvarez-Troncoso et al. 2015). Additionally, Nematoda emerges as a highly diverse group, ranking third in species richness amongst all phyla listed in our checklist. This outcome may be attributed to the interest in this phylum in ecological assessments, as nematodes have been used for this purpose for a long time (e.g. Bongers (1990); Moreno et al. (2011)). From all the sequences retrieved, based on the species of the French metropolitan freshwater macroinvertebrates' species list, 85.4% belonged to COI-5P gene (Suppl. material 1: fig. S1). This result was expected as BOLD is the main repository for COI sequences (Ratnasingham and Hebert 2007). Overall, 56% of the 2841 species listed in the French metropolitan freshwater macroinvertebrates' species list possessed at least one COI-5P genetic sequence publicly available in BOLD database (Fig. 2).

There are some exceptions with Nemerta, which have a barcode coverage of 100% and the Entoprocta that have 0% coverage (both phyla with only one species referenced). Mollusca (195 species) and Nematoda (99 species), which are the second and third most diverse phyla in terms of species in the French metropolitan freshwater macroinvertebrates' species list, have the worst barcode coverage of all phyla (41% and 5.1%, respectively). Within Arthropoda,

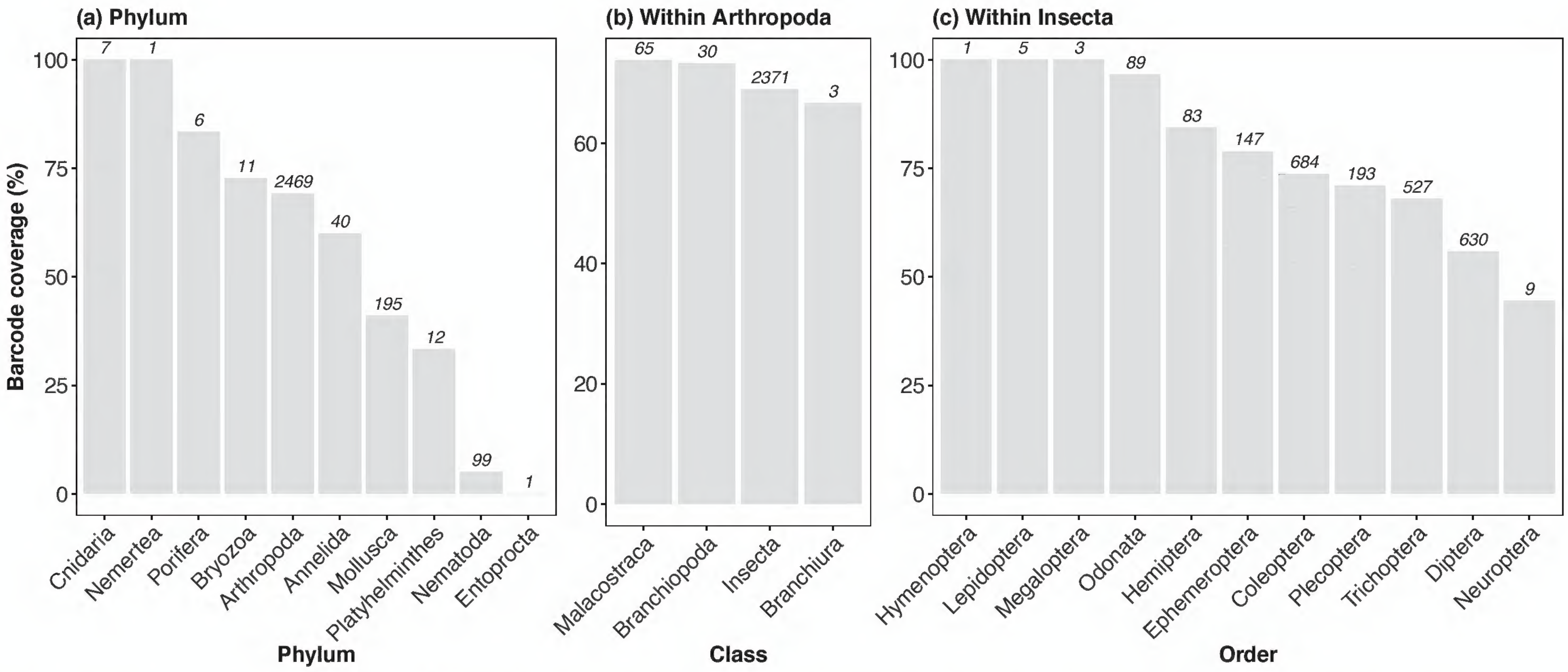


**Figure 2.** Barcoding coverage of the French metropolitan freshwater macroinvertebrates' species list. The barcoding coverage gives for each phylum, for each class within Arthropoda and for each order within Insecta the total percentage of species with available COI-5P public sequence in BOLD. The number above each bar represents the total number of species listed in the Freshwater macroinvertebrate checklist.

most classes have a barcode coverage above 60%, with the highest being in the Malacostraca (65 species, 73.8%) and the lowest in the Branchiura (3 species, 66.7%). Finally, within insects, Hymenoptera (1 specie), Lepidoptera (5 species), Megaloptera (3 species), Odonata (89 species) and Hemiptera (83 species) have the highest barcode coverage (> 80%). The three most species diverse insects' orders (Coleoptera (684 species), Diptera (630 species) and Trichoptera (527 species)) have a barcode coverage of 73.4%, 55.9% and 67.9%, respectively (Fig. 2). Compared to previous gap analyses carried out in other countries, our analysis in freshwater ecosystems in France revealed a slightly lower barcoding coverage for freshwater macroinvertebrates (56%). For instance, investigations in specific regions such as the Apulia Region of southeast Italy reported DNA barcode availability for 58% of listed aquatic Macroinvertebrate species (Specchia et al. 2020), while a study in Atlantic Iberia documented coverage for 63% of macroinvertebrates (Leite et al. 2020). Similarly, a comprehensive assessment of 4502 freshwater invertebrate species used in ecological quality assessments indicated that 60% possessed one or more barcodes (Weigand et al. 2019). Our findings of the lowest barcode coverage at the phylum level align with previous studies, which also reported very low barcode coverage for freshwater Platyhelminthes, with only three species having sequences deposited in examined databases (two in our study). The limited barcode coverage observed for this phylum may be attributed to challenges associated with amplification using standard COI primers. For instance, a study conducted across 11 rivers reported the absence of taxa detection from this phylum using three pairs of standard COI primers (Poyntz-Wright et al. 2024). A similar trend was observed for the phylum Nematoda, which also exhibits low barcode coverage and was not successfully amplified in the previously mentioned study. Furthermore, studies have demonstrated that the COI gene is an unsuitable molecular marker for free-living marine nematodes due to extensive haplotype hypervariation and frequent mitochondrial genome

recombination (Da Silva et al. 2010; Hyman et al. 2011).The low taxonomic coverage for Mollusca in our study can be attributed to the high number of DNA barcodes deposited in GenBank rather than in BOLD (Weigand et al. 2019 and references therein). Insects exhibited the highest number of available COI sequences, aligning with previous studies highlighting the over-representation of Arthropoda, particularly insects (Meglécz 2023). The barcode coverage for Insecta was 69%, aligning with other findings such as Weigand et al. (2019), who reported that 66% of monitored insect species were barcoded and Trebitz et al. (2015), who found approximately 70% representation in BOLD for Great Lakes fauna. Within Insecta, Diptera had the lowest coverage on BOLD at 55.9%, while Odonata and Hemiptera were the best covered, with over 80% of species barcoded in each group, similar to Weigand et al. (2019). The large number of DNA barcodes accessible within public databases often reflects the intensity of dedicated studies and associated barcoding projects. Certain taxonomic groups, such as Arthropoda, receive disproportionate attention, resulting in a heightened focus and increased deposition of sequences within genetic databases (Briski et al. 2011, 2016; Ardura 2019). Consequently, the absence of comprehensive reference databases may lead to false-negative outcomes (Klymus et al. 2017), while inaccuracies within these databases can engender false-positive identifications, as evidenced by instances of misreporting species presence (Port et al. 2016). Furthermore, the false assignment of sequences to closely-related species may ensue when references for the true species are absent (Schenekar et al. 2020; Couton et al. 2022). Resolving this issue necessitates the sequencing of new specimens of target taxa and their subsequent integration into reference libraries. Currently, some projects such as Biodiversity Genomics Europe (BGE) aim at producing such data, for example, DNA barcodes from 45,000 specimens of 15,000 species and metabarcoding data from environmental samples in Europe have been sequenced. In addition, some taxonomic studies released barcodes (e.g. Vuataz et al. (2024)) and ecological and evolutionary studies delivered large amounts of DNA sequences for certain European taxa (e.g. Saclier et al. (2024)).

The 1811 species, with a BIN and at least one COI-5P sequence, encompassed 3348 unique BINs, which might indicate a high rate of cryptic diversity. In one hand, this diversity is probably underestimated due to 1689 BOLD identifiers, belonging to 410 species (for which 31 are not in the 1811 species), which have no associated BINs, representing 2.4% of the total identifiers recovered (69083). This can be explained by the fact that the sequences associated with these identifiers did not meet the conditions required to be evaluated for inclusion in the BINs according to the algorithm used to delimit this cryptic diversity (Ratnasingham and Hebert 2013). On the other hand, the cryptic species estimated with the BINs could also be overestimated for the French metropolitan territory as the sequences retrieved originate from many countries (Suppl. material 1: fig. S2). Indeed, for taxa with broad geographic distributions and possibly high genetic diversity, spatial heterogeneity in the sampling may increase the splitting of divergent haplotypes attributed to a single species into multiple BINs. Furthermore, we cannot totally rule out the possibility that a BIN might also originate from contamination errors, meaning that the sequences generated and uploaded belong to bacteria or endosymbionts, for example, rather than to the intended freshwater species (Meyer and Paulay 2005; Pilgrim et al. 2021). Interestingly, 267 BINS are

associated with 410 distinct species (Suppl. material 1: table S2), which highlights that molecular techniques tend to group morphologically distinct taxa together. This may be due to the inability of the COI to differentiate between these taxa (non-discriminatory barcode) and/or to morphological over-division, based on morphological criteria that are not in agreement with the COI.

There was a significant correlation ($R^2 = 0.6$, $p < 0.01$) between the number of unique COI sequences per species and the number of BINs generated (Fig. 3).

However, some groups like Annelida or Megaloptera have a high cryptic diversity, but a low genetic diversity and, on the contrary, Platyhelminthes or Lepidoptera have a high genetic diversity, but a low cryptic diversity (Table 2). Furthermore, uncovering the number of cryptic species is tricky as the molecular delineation of species, based on a mitochondrial barcode or a small portion of it, is far from being ideal (Rubinoff et al. 2006). There are many other algorithms to evaluate this cryptic diversity with different molecular species delimitation methods (Poisson Tree Process (Kapli et al. 2017), General Mixed Yule Coalescent (Fujisawa and Barraclough 2013), Assemble Species by Automatic Partitioning (Puillandre et al. 2021), for which distinct results can be obtained for the same pool of species (Eme et al. 2018). In addition, this mitochondrial gene can also reveal a singular evolution from the nuclear genome, the so called mitonuclear discordance, calling into question the use of a single barcode for species delineation (Després 2019 and references within).

**Table 2.** Summary of species, sequences and BINs number and their ratio for each phylum, class within the arthropod phylum and order within the insect class.

| Taxonomic rank | Total number of species | Total number of sequences | Total number of unique BINS | Mean number of sequences per BINs | Mean number of BINs per species |
|---|---|---|---|---|---|
| **Phylum** | | | | | |
| Annelida | 24 | 328 | 81 | 4.0 | 3.4 |
| Arthropoda | 1708 | 33547 | 3690 | 9.1 | 2.2 |
| Bryozoa | 8 | 17 | 12 | 1.4 | 1.5 |
| Cnidaria | 7 | 245 | 34 | 7.2 | 4.9 |
| Mollusca | 80 | 2761 | 244 | 11.3 | 3.1 |
| Nematoda | 5 | 89 | 7 | 12.7 | 1.4 |
| Nemertea | 1 | 6 | 2 | 3.0 | 2.0 |
| Platyhelminthes | 4 | 104 | 5 | 20.8 | 1.3 |
| Porifera | 5 | 24 | 5 | 4.8 | 1.0 |
| **Class within Arthropoda** | | | | | |
| Branchiopoda | 22 | 815 | 90 | 9.1 | 4.1 |
| Branchiura | 2 | 25 | 5 | 5.0 | 2.5 |
| Insecta | 1636 | 28900 | 3146 | 9.2 | 1.9 |
| Malacostraca | 48 | 3807 | 449 | 8.5 | 9.4 |
| **Order within Insecta** | | | | | |
| Coleoptera | 504 | 3773 | 755 | 5.0 | 1.5 |
| Diptera | 352 | 12585 | 675 | 18.6 | 1.9 |
| Ephemeroptera | 116 | 2962 | 353 | 8.4 | 3.0 |
| Hemiptera | 70 | 781 | 113 | 6.9 | 1.6 |
| Hymenoptera | 1 | 7 | 1 | 7.0 | 1.0 |
| Lepidoptera | 5 | 79 | 8 | 9.9 | 1.6 |
| Megaloptera | 3 | 21 | 8 | 2.6 | 2.7 |
| Neuroptera | 4 | 18 | 6 | 3.0 | 1.5 |
| Odonata | 86 | 1670 | 155 | 10.8 | 1.8 |
| Plecoptera | 137 | 2105 | 300 | 7.0 | 2.2 |
| Trichoptera | 358 | 4899 | 772 | 6.3 | 2.2 |

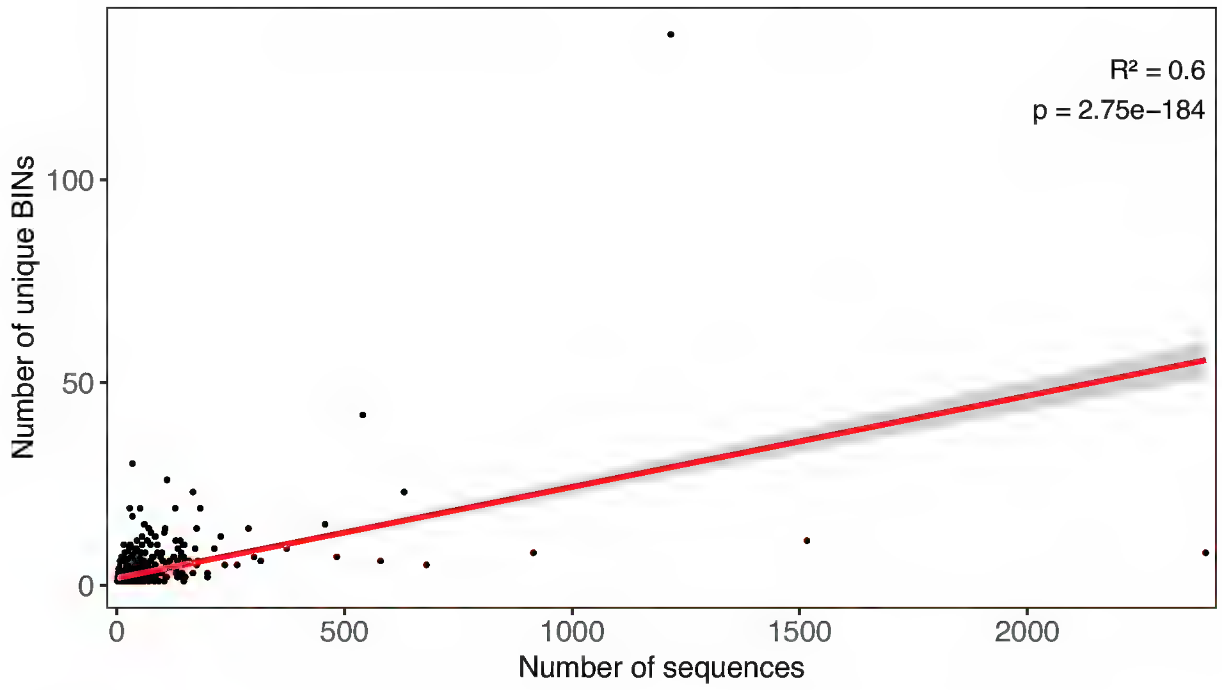


**Figure 3.** Number of BINs in relation to the number of sequences per species.

## Gap analysis of the aligned DNA library

The Fig. 4 represents the taxonomic composition of the 24 species from the French metropolitan freshwater macroinvertebrates' species list that cannot be aligned with the primer pair fwhF2/fwhR2N.

Arthropoda emerged as the predominant taxonomic groups facing alignment difficulties at the phylum level. Within insects (Arthropoda), Coleoptera and Diptera had the species with the most alignment issues. Specifically, within Malacostraca (Arthropoda), the gammarid (Gammaridae) and crayfish (Astacidea) families had the highest number of species with alignment issues (Fig. 4). Several reasons could explain these results. Firstly, although COI-5P sequences were available, some were found to lie outside the primer pairs' intended region during the alignment process, either entirely or partially. In the latter case, where the sequence in question was shorter than the desired amplified barcode, it was subsequently discarded. Another possible explanation is the poor sequence annotation, such as mislabelling (e.g. COI-3P instead of COI5-P) or misidentification of specimens, leading to incorrect species assignments. Misidentification of voucher specimens has been highlighted as a major factor contributing to erroneous records, as morphological identifications of closely-related species can be challenging. This issue has been noted by Leite et al. (2020), Pentinsaari et al. (2020) and Paz and Rinkevich (2021), emphasising its impact on subsequent species identifications using databases like BOLD. This was evident in cases when some species of insects failed to align with the primer pair, while others belonging to the same genus did. Despite BOLD being a curated database with verification procedures during sequence deposition and a reported error rate of less than 1% for Metazoan sequences at the genus level (Leray et al. 2019), problematic records may still exist, as evidenced in various studies on marine macroinvertebrates (Radulovici et al. 2021) where up to 39% of sequences were considered ambiguous. For the Decapoda species' (Malacostraca, Arthropoda), sequences that failed to align with the primers pair fwhF2/fwhR2N, comprising eight crayfish species and one crab species, their absence in the mock community during primer design could account for this discrepancy (Elbrecht and Leese 2015; Elbrecht et al. 2017). Although Gammarids (Amphipoda, Arthropoda) were included in primer design, the

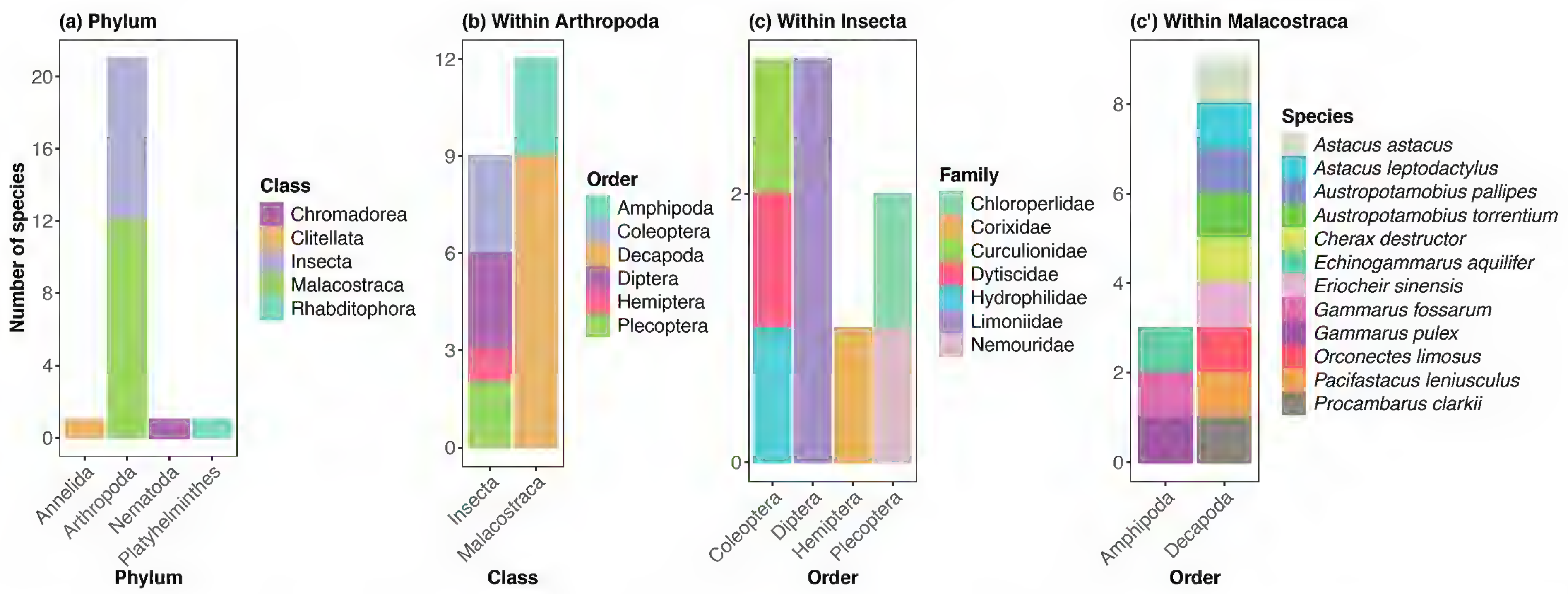


**Figure 4.** Number of species associated with COI-5P genetic sequences that cannot be aligned with fwh2 primer pair at the phylum rank (**a**), within the Arthropoda phylum (**b**) and within the Insecta and Malacostraca class (**c, c'**).

failure of some species' sequences to align with the primers could be due to their vast diversity within aquatic environments (Horton et al. 2023). Molecular studies on Amphipoda have revealed extensive species diversity and the presence of cryptic species complexes (Jażdżewska and Mamos 2019), suggesting a high genetic variability within species (Lefébure et al. 2006) that may not be compatible with the primer pair. Furthermore, Vamos et al. (2017) indicated that their analysis of the efficiency of the primers pair displayed higher penalty scores for certain taxa of Turbellaria, Mollusca, Trichoptera and Isopoda, indicating potential under-representation due to primer bias. This aligns with findings from other studies suggesting preferential detection of taxonomic groups by different markers and primers (Leduc et al. 2019). Consequently, the incorporation of multiple genetic regions in DNA metabarcoding to ensure the broadest possible taxonomic detection may prove to be a good solution (Duarte et al. 2021). This could be particularly important to monitor the noble crayfish, *Astacus astacus* and the amphipod *Gammarus roeselii* which cannot be detected with fwhF2/fwhR2N primers pair (Fig. 4), whereas they are the most frequently monitored species of the malacostracans in Europe (Weigand et al. 2019). Analysis of the number of haplotypes relative to the number of species available within different phyla (Fig. 5) provides important information about the potential to detect taxa in natural samples.

In our reference library, for most phyla, the majority of species have less than five haplotypes, except cnidarians, which have the majority of their species with 5–25 haplotypes. This pattern of low haplotypes availability could be due to a lack of samples; therefore, too few taxa are sequenced to detect genetic variability. Only three species presented more than 100 unique barcodes: one trichopteran (*Agraylea multipunctata*), one isopod (*Asellus aquaticus*) and one gasteropod (*Physella acuta*). These findings are consistent with the observations of Trebitz et al. (2015), suggesting a noteworthy exception to the prevailing low barcoding rate for invertebrates, with some species being exceptionally well genetically represented. This phenomenon may be attributed to the scientific significance and ecological relevance of these species, such as being acknowledged as a reliable indicator. Indeed, *Asellus aquaticus* is a common species monitored in European countries

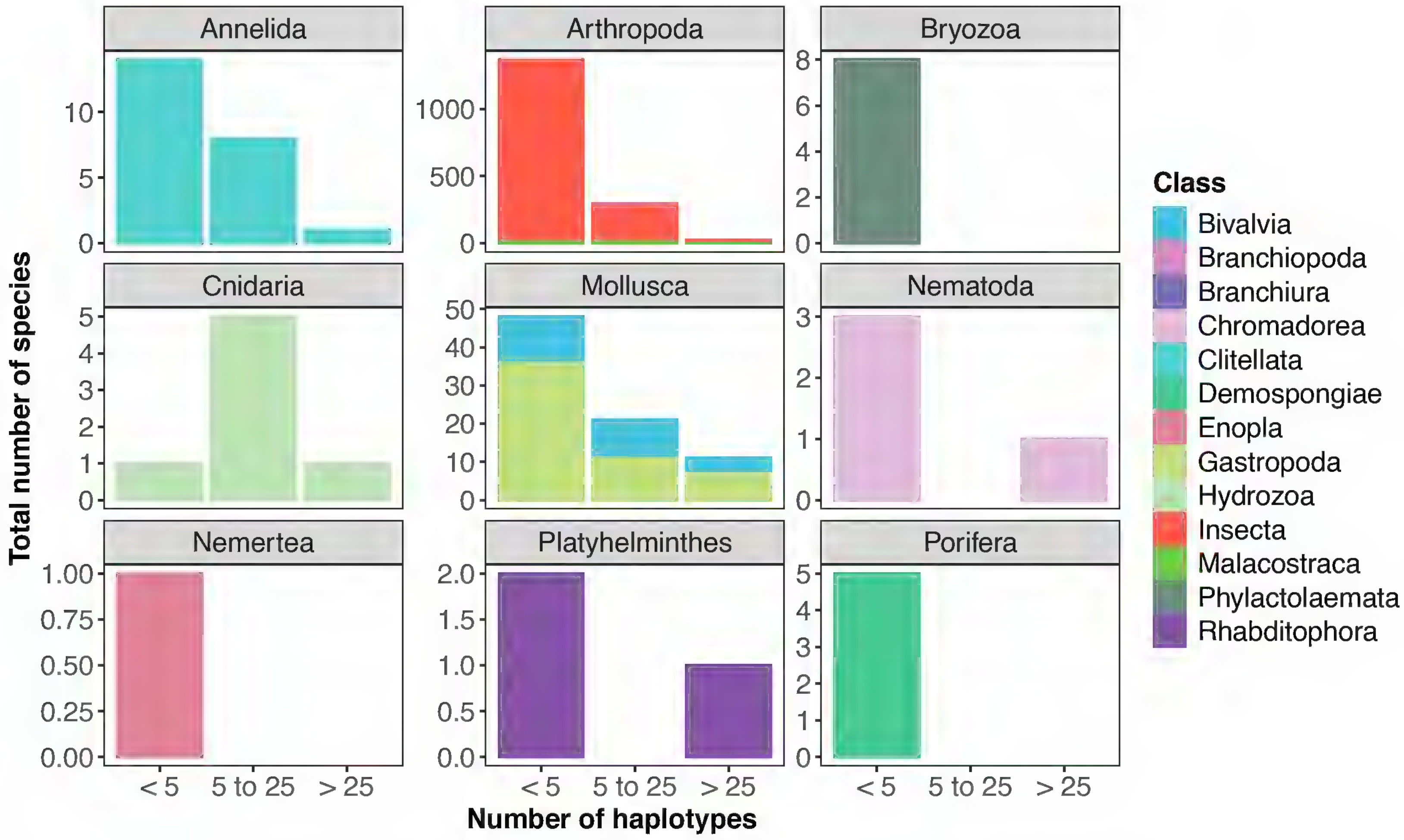


**Figure 5.** Taxonomic composition at the class rank for the species from the aligned DNA library. Each box at the phylum level represents the distribution of COI-5P haplotypes relative to the number of species.

(Weigand et al. 2019), known as a bioindicator for metal pollutants detection (e.g. O'Callaghan et al. (2019)). *Physella acuta* is also intensively studied as it is an invasive aquatic Gasteropoda with worldwide distribution (Banha et al. 2014; Vinarski 2017). The analysis on the ability of the fwhF2/fwhR2N primers to discriminate each species by a unique sequence demonstrated that 57 identical sequences were shared by 116 distinct species (Suppl. material 1: table S3). Amongst those sequences, 10 were shared by two or three different genera and 47 were shared by two or three different species. These results could be explained by the fact that some taxa could have been mislabelled or that, for this length of barcode (205 bp), no genetic variability is found between two related taxa; therefore, this barcode is not suitable to decipher those species. Several authors showed the necessity to have multiple sequences for each species to cover correctly their haplotypic diversity (Leite et al. 2020; Keck et al. 2023). This imperative arises from the recognition that the absence of intraspecific variants can pose significant challenges. In instances where a single sequence is available for a species exhibiting high genetic variability, the accurate identification of all haplotypes may be compromised. Moreover, inadequate representation of closely-related species within reference libraries can lead to the erroneous assignment of multiple species to a single taxonomic entity (e.g. Jackman et al. (2021)), potentially resulting in erroneous assessments of species diversity.

## Conclusions

Our study underscores a widespread absence of reference barcodes for numerous extant invertebrate species. Although barcoding offers advantages over morphological identification in biomonitoring, existing gaps in barcode libraries may hinder their effectiveness (Duarte et al. 2020; Feio et al. 2020;

Hestetun et al. 2020; Vieira et al. 2021). Substantial efforts are necessary to sequence new individuals for species absent from the reference barcoding library and for species with a low number of sequences. This will enable a more efficient and reliable species identification and biodiversity assessment, since the effectiveness and reliability of DNA barcoding identification are linked to the thoroughness of taxonomic curation and completeness of the reference barcoding libraries (Geiger et al. 2021; Keck et al. 2023). Nevertheless, our work has established a reference barcoding library for freshwater macroinvertebrates at the French level. This database can be used for biodiversity studies employing environmental DNA with short COI primers (fwhF2/fwhR2n) designed for samples containing partially degraded DNA. Furthermore, we have provided insights into the various biases associated with the use of this library that future users will have to consider to interpret their results. Looking ahead, an important future development for the fwhF2/fwhRNn reference library would be to assign a confidence level to the identification of each taxon (e.g. species). This confidence level could be determined, based on several factors, including the number of reference sequences available (with higher sequence counts providing greater confidence), geographical coverage (broader coverage being preferable), species delimitation (with monophyletic groups offering more robust identification compared to paraphyletic ones) and the availability of metadata associated with each sequence (e.g. sampling date and location, habitat, sequence quality). These criteria align with recommendations proposed by Fontes et al. (2021).

## Acknowledgements

We would also like to thank Tristan Lefébure and Maylïs Gauthier from LEHNA (CNRS) for their insights into this study. We would also like to thank the reviewers for correcting and improving this study. The authors have declared that no competing interests exist.

## Additional information

### Conflict of interest

The authors have declared that no competing interests exist.

### Ethical statement

No ethical statement was reported.

### Funding

This study was made possible thanks to a funding of the Carnot Eau & Environnement institute, a funding of the INRAE department AQUA and a funding of the pôle INRAE-OFB ECLA (ECosystèmes LAcustres).

### Author contributions

Conceptualization: PG, FR. Data curation: PG. Formal analysis: DE, PG. Investigation: DE, PG. Methodology: PG, DE. Supervision: ID, FR. Validation: ID, DE, FR. Visualization: ID, DE, FR. Writing - original draft: PG. Writing - review and editing: DE, ID, FR.

## Author ORCIDs

Paula Gauvin https://orcid.org/0000-0002-9268-1856
David Eme https://orcid.org/0000-0001-8790-0412
Frédéric Rimet https://orcid.org/0000-0002-5514-869X


## Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

## References


Abad D, Albaina A, Aguirre M, Laza-Martínez A, Uriarte I, Iriarte A, Villate F, Estonba A (2016) Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Marine Biology 163(7): 149. https://doi.org/10.1007/s00227-016-2920-0

Álvarez-Troncoso R, Benetti CJ, Sarr AB, Pérez-Bilbao A, Garrido J (2015) Impacts of hydroelectric power stations on Trichoptera assemblages in four rivers in NW Spain. Limnologica 53: 35–41. https://doi.org/10.1016/j.limno.2015.05.001

Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology 27(20): 3968–3975. https://doi.org/10.1111/mec.14844

Ardura A (2019) Species-specific markers for early detection of marine invertebrate invaders through eDNA methods: Gaps and priorities in GenBank as database example. Journal for Nature Conservation 47: 51–57. https://doi.org/10.1016/j.jnc.2018.11.005

Banha F, Marques M, Anastácio PM (2014) Dispersal of two freshwater invasive macroinvertebrates, *Procambarus clarkii* and *Physella acuta*, by off-road vehicles. Aquatic Conservation 24(5): 582–591. https://doi.org/10.1002/aqc.2453

Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM (2014) Environmental Conditions Influence eDNA Persistence in Aquatic Systems. Environmental Science & Technology 48(3): 1819–1827. https://doi.org/10.1021/es404734p

Bean TP, Greenwood N, Beckett R, Biermann L, Bignell JP, Brant JL, Copp GH, Devlin MJ, Dye S, Feist SW, Fernand L, Foden D, Hyder K, Jenkins CM, Van Der Kooij J, Kröger S, Kupschus S, Leech C, Leonard KS, Lynam CP, Lyons BP, Maes T, Nicolaus EEM, Malcolm SJ, McIlwaine P, Merchant ND, Paltriguera L, Pearce DJ, Pitois SG, Stebbing PD, Townhill B, Ware S, Williams O, Righton D (2017) A review of the tools used for marine monitoring in the UK: Combining historic and contemporary methods with modeling and socioeconomics to fulfill legislative needs and scientific ambitions. Frontiers in Marine Science 4: 263. https://doi.org/10.3389/fmars.2017.00263

Blackman R, Mächler E, Altermatt F, Arnold A, Beja P, Boets P, Egeter B, Elbrecht V, Filipe AF, Jones J, Macher J, Majaneva M, Martins F, Múrria C, Meissner K, Pawlowski J, Schmidt Yáñez P, Zizka V, Leese F, Price B, Deiner K (2019) Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. Metabarcoding and Metagenomics 3: e34735. https://doi.org/10.3897/mbmg.3.34735

Blackman RC, Walser J, Rüber L, Brantschen J, Villalba S, Brodersen J, Seehausen O, Altermatt F (2023) General principles for assignments of communities from eDNA : Open versus closed taxonomic databases. Environmental DNA 5(2): 326–342. https://doi.org/10.1002/edn3.382

Bonada N, Prat N, Resh VH, Statzner B (2006) Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. Annual Review of Entomology 51(1): 495–523. https://doi.org/10.1146/annurev.ento.51.110104.151124

Bongers T (1990) The maturity index: An ecological measure of environmental disturbance based on nematode species composition. Oecologia 83(1): 14–19. https://doi.org/10.1007/BF00324627

Borgwardt F, Robinson L, Trauner D, Teixeira H, Nogueira AJA, Lillebø AI, Piet G, Kuemmerlen M, O'Higgins T, McDonald H, Arevalo-Torres J, Barbosa AL, Iglesias-Campos A, Hein T, Culhane F (2019) Exploring variability in environmental impact risk from human activities across aquatic ecosystems. The Science of the Total Environment 652: 1396–1408. https://doi.org/10.1016/j.scitotenv.2018.10.339

Briski E, Cristescu ME, Bailey SA, MacIsaac HJ (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. Biological Invasions 13(6): 1325–1340. https://doi.org/10.1007/s10530-010-9892-7

Briski E, Ghabooli S, Bailey SA, MacIsaac HJ (2016) Are genetic databases sufficiently populated to detect non-indigenous species? Biological Invasions 18(7): 1911–1922. https://doi.org/10.1007/s10530-016-1134-1

Bush A, Compson ZG, Monk WA, Porter TM, Steeves R, Emilson E, Gagne N, Hajibabaei M, Roy M, Baird DJ (2019) Studying Ecosystems With DNA Metabarcoding: Lessons From Biomonitoring of Aquatic Macroinvertebrates. Frontiers in Ecology and Evolution 7: 434. https://doi.org/10.3389/fevo.2019.00434

Caesar RM, Sörensson M, Cognato AI (2006) Integrating DNA data and traditional taxonomy to streamline biodiversity assessment: An example from edaphic beetles in the Klamath ecoregion, California, USA. Diversity & Distributions 12(5): 483–489. https://doi.org/10.1111/j.1366-9516.2006.00237.x

Carraro L, Altermatt F (2024) eDITH: an R-package to spatially project eDNA-based biodiversity across river networks with minimal prior information. Ecology. preprint https://doi.org/10.1101/2024.01.16.575835

Choudhary A, Ahi J (2015) Biodiversity of freshwater insects. RE:view.

Couton M, Lévêque L, Daguin-Thiébaut C, Comtet T, Viard F (2022) Water eDNA metabarcoding is effective in detecting non-native species in marinas, but detection errors still hinder its use for passive monitoring. Biofouling 38(4): 367–383. https://doi.org/10.1080/08927014.2022.2075739

Csabai Z, Čiamporová-Zaťovičová Z, Boda P, Čiampor Jr F (2023) 50%, not great, not terrible: Pan-European gap-analysis shows the real status of the DNA barcode reference libraries in two aquatic invertebrate groups and points the way ahead. The Science of the Total Environment 863: 160922. https://doi.org/10.1016/j.scitotenv.2022.160922

Da Silva NRR, Da Silva MC, Fonseca Genevois V, Esteves AM, Ley PD, Decraemer W, Rieger TT, Dos Santos Correia MT (2010) Marine nematode taxonomy in the age of DNA: The present and future of molecular tools to assess their biodiversity. Nematology 12(5): 661–672. https://doi.org/10.1163/138855410X500073

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–5895. https://doi.org/10.1111/mec.14350

Després L (2019) One, two or more species? Mitonuclear discordance and species delimitation. Molecular Ecology 28(17): 3845–3847. https://doi.org/10.1111/mec.15211

Dijkstra K-DB, Monaghan MT, Pauls SU (2014) Freshwater Biodiversity and Aquatic Insect Diversification. Annual Review of Entomology 59(1): 143–163. https://doi.org/10.1146/annurev-ento-011613-161958

Duarte S, Vieira PE, Costa FO (2020) Assessment of species gaps in DNA barcode libraries of non-indigenous species (NIS) occurring in European coastal regions. Metabarcoding and Metagenomics 4: e55162. https://doi.org/10.3897/mbmg.4.55162

Duarte S, Leite B, Feio M, Costa F, Filipe A (2021) Integration of DNA-Based Approaches in Aquatic Ecological Assessment Using Benthic Macroinvertebrates. Water (Basel) 13(3): 331. https://doi.org/10.3390/w13030331

Elbrecht V, Leese F (2015) Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. PLoS ONE 10: e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J-N, Coissac E, Boyer F, Leese F (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4: e1966. https://doi.org/10.7717/peerj.1966

Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution 8: 1265–1275. https://doi.org/10.1111/2041-210X.12789

Eme D, Zagmajster M, Delić T, Fišer C, Flot J, Konecny-Dupré L, Pálsson S, Stoch F, Zakšek V, Douady CJ, Malard F (2018) Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. Ecography 41(2): 424–436. https://doi.org/10.1111/ecog.02683

Feio MJ, Garcia-Reventós A, Ardura A, Calapez AR, Pujante AM, Mortágua A, Múrria C, Diaz-de-Quijano D, Filipe AF (2020) Advances in the use of molecular tools in ecological and biodiversity assessment of aquatic ecosystems. Limnetica 39(1): 419–440. https://doi.org/10.23818/limn.39.27

Ficetola GF, Boyer F, Valentini A, Bonin A, Meyer A, Dejean T, Gaboriaud C, Usseglio-Polatera P, Taberlet P (2021) Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. Molecular Ecology 30(13): 3189–3202. https://doi.org/10.1111/mec.15632

Fontes JT, Vieira PE, Ekrem T, Soares P, Costa FO (2021) BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. Molecular Ecology Resources 21(2): 573–583. https://doi.org/10.1111/1755-0998.13262

Fujisawa T, Barraclough TG (2013) Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. Systematic Biology 62(5): 707–724. https://doi.org/10.1093/sysbio/syt033

Geiger M, Koblmüller S, Assandri G, Chovanec A, Ekrem T, Fischer I, Galimberti A, Grabowski M, Haring E, Hausmann A, Hendrich L, Koch S, Mamos T, Rothe U, Rulik B, Rewicz T, Sittenthaler M, Stur E, Tończyk G, Zangl L, Moriniere J (2021) Coverage and quality of DNA barcode references for Central and Northern European Odonata. PeerJ 9: e11192. https://doi.org/10.7717/peerj.11192

Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M (2015) Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. PLoS ONE 10: e0138432. https://doi.org/10.1371/journal.pone.0138432

Grosberg RK, Vermeij GJ, Wainwright PC (2012) Biodiversity in water and on land. Current Biology 22(21): R900–R903. https://doi.org/10.1016/j.cub.2012.09.050

Hebert PDN, Ratnasingham S, De Waard JR (2003a) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences 270. https://doi.org/10.1098/rsbl.2003.0025

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003b) Biological identifications through DNA barcodes. Proceedings. Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of Birds through DNA Barcodes. PLoS Biology 2: e312. https://doi.org/10.1371/journal.pbio.0020312

Hering D, Meier C, Rawer-Jost C, Feld CK, Biss R, Zenker A, Sundermann A, Lohse S, Böhmer J (2004) Assessing streams in Germany with benthic invertebrates: Selection of candidate metrics. Limnologica 34(4): 398–415. https://doi.org/10.1016/S0075-9511(04)80009-4

Hering D, Johnson RK, Kramm S, Schmutz S, Szoszkiewicz K, Verdonschot PFM (2006) Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. Freshwater Biology 51(9): 1757–1785. https://doi.org/10.1111/j.1365-2427.2006.01610.x

Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hestetun JT, Bye-Ingebrigtsen E, Nilsson RH, Glover AG, Johansen P-O, Dahlgren TG (2020) Significant taxon sampling gaps in DNA databases limit the operational use of marine macrofauna metabarcoding. Marine Biodiversity 50(5): 70. https://doi.org/10.1007/s12526-020-01093-5

Holman LE, De Bruyn M, Creer S, Carvalho G, Robidart J, Rius M (2019) Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. Scientific Reports 9(1): 11559. https://doi.org/10.1038/s41598-019-47899-7

Horton T, De Broyer C, Bellan-Santini D, Coleman CO, Copilaș-Ciocianu D, Corbari L, Daneliya ME, Dauvin J-C, Decock W, Fanini L, Fišer C, Gasca R, Grabowski M, Guerra-García JM, Hendrycks EA, Hughes LE, Jaume D, Kim Y-H, King RA, Lo Brutto S, Lörz A-N, Mamos T, Serejo CS, Senna AR, Souza-Filho JF, Tandberg AHS, Thurston MH, Vader W, Väinölä R, Valls Domedel G, Vandepitte L, Vanhoorne B, Vonk R, White KN, Zeidler W (2023) The World Amphipoda Database: History and progress. Records of the Australian Museum 75(4): 329–342. https://doi.org/10.3853/j.2201-4349.75.2023.1875

Hyman BC, Lewis SC, Tang S, Wu Z (2011) Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. Genetica 139(5): 611–615. https://doi.org/10.1007/s10709-010-9531-3

Jäch MA, Balke M (2008) Global diversity of water beetles (Coleoptera) in freshwater. Hydrobiologia 595(1): 419–442. https://doi.org/10.1007/s10750-007-9117-y

Jackman JM, Benvenuto C, Coscia I, Oliveira Carvalho C, Ready JS, Boubli JP, Magnusson WE, McDevitt AD, Guimarães Sales N (2021) eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. Environmental DNA 3(4): 837–849. https://doi.org/10.1002/edn3.191

Jażdżewska AM, Mamos T (2019) High species richness of Northwest Pacific deep-sea amphipods revealed through DNA barcoding. Progress in Oceanography 178: 102184. https://doi.org/10.1016/j.pocean.2019.102184

Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T (2017) Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33: 1630–1638. https://doi.org/10.1093/bioinformatics/btx025

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Klymus KE, Marshall NT, Stepien CA (2017) Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. PLoS ONE 12: e0177643. https://doi.org/10.1371/journal.pone.0177643

Kozlov AM, Zhang J, Yilmaz P, Glöckner FO, Stamatakis A (2016) Phylogeny-aware identification and correction of taxonomically mislabeled sequences. Nucleic Acids Research 44(11): 5022–5033. https://doi.org/10.1093/nar/gkw396

Leduc N, Lacoursière-Roussel A, Howland KL, Archambault P, Sevellec M, Normandeau E, Dispas A, Winkler G, McKindsey CW, Simard N, Bernatchez L (2019) Comparing eDNA metabarcoding and species collection for documenting Arctic metazoan biodiversity. Environmental DNA 1(4): 342–358. https://doi.org/10.1002/edn3.35

Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J, Piggott J, Rimet F, Steinke D, Taberlet P, Weigand A, Abarenkov K, Beja P, Bervoets L, Björnsdóttir S, Boets P, Boggero A, Bones A, Borja Á, Bruce K, Bursić V, Carlsson J, Čiampor F, Čiamporová-Zatovičová Z, Coissac E, Costa F, Costache M, Creer S, Csabai Z, Deiner K, DelValls Á, Drakare S, Duarte S, Eleršek T, Fazi S, Fišer C, Flot J-F, Fonseca V, Fontaneto D, Grabowski M, Graf W, Guðbrandsson J, Hellström M, Hershkovitz Y, Hollingsworth P, Japoshvili B, Jones J, Kahlert M, Kalamujic Stroil B, Kasapidis P, Kelly M, Kelly-Quinn M, Keskin E, Kõljalg U, Ljubešić Z, Maček I, Mächler E, Mahon A, Marečková M, Mejdandzic M, Mircheva G, Montagna M, Moritz C, Mulk V, Naumoski A, Navodaru I, Padisák J, Pálsson S, Panksep K, Penev L, Petrusek A, Pfannkuchen M, Primmer C, Rinkevich B, Rotter A, Schmidt-Kloiber A, Segurado P, Speksnijder A, Stoev P, Strand M, Šulčius S, Sundberg P, Traugott M, Tsigenopoulos C, Turon X, Valentini A, Van Der Hoorn B, Várbíró G, Vasquez Hadjilyra M, Viguri J, Vitonytė I, Vogler A, Vrålstad T, Wägele W, Wenne R, Winding A, Woodward G, Zegura B, Zimmermann J (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes 2: e11321. https://doi.org/10.3897/rio.2.e11321

Leese F, Bouchez A, Abarenkov K, Altermatt F, Borja Á, Bruce K, Ekrem T, Čiampor F, Čiamporová-Zaťovičová Z, Costa FO, Duarte S, Elbrecht V, Fontaneto D, Franc A, Geiger MF, Hering D, Kahlert M, Kalamujić Stroil B, Kelly M, Keskin E, Liska I, Mergen P, Meissner K, Pawlowski J, Penev L, Reyjol Y, Rotter A, Steinke D, Van Der Wal B, Vitecek S, Zimmermann J, Weigand AM (2018) Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action. Advances in Ecological Research, Elsevier, 63–99. https://doi.org/10.1016/bs.aecr.2018.01.001

Lefébure T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. Molecular Phylogenetics and Evolution 40(2): 435–447. https://doi.org/10.1016/j.ympev.2006.03.014

Leite BR, Vieira PE, Teixeira MAL, Lobo-Arteaga J, Hollatz C, Borges LMS, Duarte S, Troncoso JS, Costa FO (2020) Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia. Regional Studies in Marine Science 36: 101307. https://doi.org/10.1016/j.rsma.2020.101307

Leray M, Knowlton N, Ho S-L, Nguyen BN, Machida RJ (2019) GenBank is a reliable resource for 21st century biodiversity research. Proceedings of the National Academy of Sciences of the United States of America 116(45): 22651–22656. https://doi.org/10.1073/pnas.1911714116

Li J, Lawson Handley LJ, Harper LR, Brys R, Watson HV, Di Muri C, Zhang X, Hänfling B (2019) Limited dispersion and quick degradation of environmental DNA in fish ponds inferred by metabarcoding. Environmental DNA 1(3): 238–250. https://doi.org/10.1002/edn3.24

Locatelli NS, McIntyre PB, Therkildsen NO, Baetscher DS (2020) GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. Proceedings of the National Academy of Sciences of the United States of America 117(51): 32211–32212. https://doi.org/10.1073/pnas.2007421117

May ML (2019) Dispersal by Aquatic Insects. In: Del-Claro K, Guillermo R (Eds) Aquatic Insects. Springer International Publishing, Cham, 35–73. https://doi.org/10.1007/978-3-030-16327-3_3

Meglécz E (2023) COINR and MKCOINR : Building and customizing a nonredundant barcoding reference database from BOLD and NCBI using a semi-automated pipeline. Molecular Ecology Resources 23: 933–945. https://doi.org/10.1111/1755-0998.13756

Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9(1): 214. https://doi.org/10.1186/1471-2164-9-214

Meyer CP, Paulay G (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biology 3: e422. https://doi.org/10.1371/journal.pbio.0030422

Mondy CP, Villeneuve B, Archaimbault V, Usseglio-Polatera P (2012) A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach. Ecological Indicators 18: 452–467. https://doi.org/10.1016/j.ecolind.2011.12.013

Moreno M, Semprucci F, Vezzulli L, Balsamo M, Fabiano M, Albertelli G (2011) The use of nematodes in assessing ecological quality status in the Mediterranean coastal ecosystems. Ecological Indicators 11(2): 328–336. https://doi.org/10.1016/j.ecolind.2010.05.011

Mugnai F, Costantini F, Chenuil A, Leduc M, Gutiérrez Ortega JM, Meglécz E (2023) Be positive: Customized reference databases and new, local barcodes balance false taxonomic assignments in metabarcoding studies. PeerJ 11: e14616. https://doi.org/10.7717/peerj.14616

O'Callaghan I, Harrison S, Fitzpatrick D, Sullivan T (2019) The freshwater isopod Asellus aquaticus as a model biomonitor of environmental pollution: A review. Chemosphere 235: 498–509. https://doi.org/10.1016/j.chemosphere.2019.06.217

Oliveira LM, Knebelsberger T, Landi M, Soares P, Raupach MJ, Costa FO (2016) Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. Journal of Fish Biology 89(6): 2741–2754. https://doi.org/10.1111/jfb.13169

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, Van Der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquín J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R,

Vasselon V, Vitecek S, Zimmerman J, Weigand A, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. The Science of the Total Environment 637–638: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Paz G, Rinkevich B (2021) Gap analysis of DNA barcoding in ERMS reference libraries for ascidians and cnidarians. Environmental Sciences Europe 33(1): 4. https://doi.org/10.1186/s12302-020-00449-9

Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN (2020) BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? PLoS ONE 15: e0231814. https://doi.org/10.1371/journal.pone.0231814

Pilgrim J, Thongprem P, Davison HR, Siozios S, Baylis M, Zakharov EV, Ratnasingham S, deWaard JR, Macadam CR, Smith MA, Hurst GDD (2021) Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis. GigaScience 10(3): giab021. https://doi.org/10.1093/gigascience/giab021

Pinder LCV (1986) Biology of Freshwater Chironomidae. Annual Review of Entomology 31(1): 1–23. https://doi.org/10.1146/annurev.en.31.010186.000245

Pochon X, Zaiko A, Hopkins GA, Banks JC, Wood SA (2015) Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: An assessment of different sampling devices. Biofouling 31(3): 241–251. https://doi.org/10.1080/08927014.2015.1028923

Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP (2016) Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. Molecular Ecology 25: 527–541. https://doi.org/10.1111/mec.13481

Poyntz-Wright IP, Harrison XA, Pedersen S, Tyler CR (2024) Effectiveness of eDNA for monitoring riverine macroinvertebrates. The Science of the Total Environment 941: 173621. https://doi.org/10.1016/j.scitotenv.2024.173621

Puillandre N, Brouillet S, Achaz G (2021) ASAP: Assemble species by automatic partitioning. Molecular Ecology Resources 21(2): 609–620. https://doi.org/10.1111/1755-0998.13281

Radulovici AE, Vieira PE, Duarte S, Teixeira MAL, Borges LMS, Deagle BE, Majaneva S, Redmond N, Schultz JA, Costa FO (2021) Revision and annotation of DNA barcode records for marine invertebrates: Report of the 8th iBOL conference hackathon. Metabarcoding and Metagenomics 5: e67862. https://doi.org/10.3897/mbmg.5.67862

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7(3): 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8: e66213. https://doi.org/10.1371/journal.pone.0066213

Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019) Emerging threats and persistent conservation challenges for freshwater biodiversity. Biological Reviews of the Cambridge Philosophical Society 94(3): 849–873. https://doi.org/10.1111/brv.12480

Richardson RT, Bengtsson-Palme J, Gardiner MM, Johnson RM (2018) A reference cytochrome c oxidase subunit I database curated for hierarchical classification of arthropod metabarcoding data. PeerJ 6: e5126. https://doi.org/10.7717/peerj.5126

Rivera SF, Vasselon V, Mary N, Monnier O, Rimet F, Bouchez A (2021) Exploring the capacity of aquatic biofilms to act as environmental DNA samplers: Test on

macroinvertebrate communities in rivers. Science of The Total Environment 763: 144208. https://doi.org/10.1016/j.scitotenv.2020.144208

Rodriguez-Ezpeleta N, Morissette O, Bean CW, Manu S, Banerjee P, Lacoursière-Roussel A, Beng KC, Alter SE, Roger F, Holman LE, Stewart KA, Monaghan MT, Mauvisseau Q, Mirimin L, Wangensteen OS, Antognazza CM, Helyar SJ, Boer H, Monchamp M, Nijland R, Abbott CL, Doi H, Barnes MA, Leray M, Hablützel PI, Deiner K (2021) Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on "Environmental DNA: What's behind the term?" by Pawlowski et al., (2020). Molecular Ecology 30(19): 4601–4605. https://doi.org/10.1111/mec.15942

Rubinoff D, Cameron S, Will K (2006) A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification. The Journal of Heredity 97(6): 581–594. https://doi.org/10.1093/jhered/esl036

Saclier N, Duchemin L, Konecny-Dupré L, Grison P, Eme D, Martin C, Callou C, Lefébure T, François C, Issartel C, Lewis JJ, Stoch F, Sket B, Gottstein S, Delić T, Zagmajster M, Grabowski M, Weber D, Reboleira ASPS, Palatov D, Paragamian K, Knight LRFD, Michel G, Lefebvre F, Hosseini MM, Camacho AI, De Bikuña BG, Taleb A, Belaidi N, Tuekam Kayo RP, Galassi DMP, Moldovan OT, Douady CJ, Malard F (2024) A collaborative backbone resource for comparative studies of subterranean evolution: The World Asellidae database. Molecular Ecology Resources 24(1): e13882. https://doi.org/10.1111/1755-0998.13882

Sakata MK, Yamamoto S, Gotoh RO, Miya M, Yamanaka H, Minamoto T (2020) Sedimentary eDNA provides different information on timescale and fish species composition compared with aqueous eDNA. Environmental DNA 2(4): 505–518. https://doi.org/10.1002/edn3.75

Schenekar T, Schletterer M, Lecaudey LA, Weiss SJ (2020) Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. River Research and Applications 36(7): 1004–1013. https://doi.org/10.1002/rra.3610

Seymour M, Durance I, Cosby BJ, Ransom-Jones E, Deiner K, Ormerod SJ, Colbourne JK, Wilgar G, Carvalho GR, De Bruyn M, Edwards F, Emmett BA, Bik HM, Creer S (2018) Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. Communications Biology 1(1): 4. https://doi.org/10.1038/s42003-017-0005-3

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research: Next-generation sequencing for environmental DNA. Molecular Ecology 21(8): 1794–1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x

Short AEZ (2018) Systematics of aquatic beetles (Coleoptera): Current state and future directions. Systematic Entomology 43(1): 1–18. https://doi.org/10.1111/syen.12270

Specchia V, Tzafesta E, Marini G, Scarcella S, D'Attis S, Pinna M (2020) Gap Analysis for DNA Barcode Reference Libraries for Aquatic Macroinvertebrate Species in the Apulia Region (Southeast of Italy). Journal of Marine Science and Engineering 8(7): 538. https://doi.org/10.3390/jmse8070538

Steinegger M, Salzberg SL (2020) Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. Genome Biology 21(1): 115. https://doi.org/10.1186/s13059-020-02023-1

Sweeney BW, Battle JM, Jackson JK, Dapkey T (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? Journal of the North American Benthological Society 30(1): 195–216. https://doi.org/10.1899/10-016.1

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA: ENVIRONMENTAL DNA. Molecular Ecology 21(8): 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Tachet H, Richoux P, Bournaud M, Usseglio-Polatera P (2000) Invertébrés d'eau douce - systématique, biologie, écologie - CNRS Editions. CNRS. https://www.cnrseditions.fr/catalogue/ecologie-environnement-sciences-de-la-terre/invertebres-d-eau-douce-henri-tachet/ [July 15, 2024]

Trebitz AS, Hoffman JC, Grant GW, Billehus TM, Pilgrim EM (2015) Potential for DNA-based identification of Great Lakes fauna: Match and mismatch between taxa inventories and DNA barcode libraries. Scientific Reports 5(1): 12162. https://doi.org/10.1038/srep12162

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25(4): 929–942. https://doi.org/10.1111/mec.13428

Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. https://doi.org/10.3897/mbmg.1.14625

Van Der Lee GH, Polling M, Van Der Laan I, Kodde L, Verdonschot RCM (2024) From DNA to diagnostics: A case study using macroinvertebrate metabarcoding to assess the effectiveness of restoration measures in a Dutch stream. The Science of the Total Environment 923: 171413. https://doi.org/10.1016/j.scitotenv.2024.171413

Vieira PE, Lavrador AS, Parente MI, Parretti P, Costa AC, Costa FO, Duarte S (2021) Gaps in DNA sequence libraries for Macaronesian marine macroinvertebrates imply decades till completion and robust monitoring. Diversity & Distributions 27(10): 2003–2015. https://doi.org/10.1111/ddi.13305

Vinarski MV (2017) The history of an invasion: Phases of the explosive spread of the physid snail Physella acuta through Europe, Transcaucasia and Central Asia. Biological Invasions 19(4): 1299–1314. https://doi.org/10.1007/s10530-016-1339-3

Vuataz L, Reding J-P, Reding A, Roesti C, Stoffel C, Vinçon G, Gattolliat J-L (2024) A comprehensive DNA barcoding reference database for Plecoptera of Switzerland. Scientific Reports 14(1): 6322. https://doi.org/10.1038/s41598-024-56930-5

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra K-DB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, Van Der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. The Science of the Total Environment 678: 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Yeates DK, Wiegmann BM (1999) Congruence and controversy: Toward a higher-level phylogeny of diptera. Annual Review of Entomology 44(1): 397–428. https://doi.org/10.1146/annurev.ento.44.1.397

Young HS, McCauley DJ, Galetti M, Dirzo R (2016) Patterns, Causes, and Consequences of Anthropocene Defaunation. Annual Review of Ecology, Evolution, and Systematics 47(1): 333–358. https://doi.org/10.1146/annurev-ecolsys-112414-054142

## Supplementary material 1

### Taxonomic harmonization of the French metropolitan freshwater macroinvertebrates' species list

Authors: Paula Gauvin, David Eme, Isabelle Domaizon, Frédéric Rimet

Data type: xlsx

Explanation note: This table lists the taxa requiring taxonomic harmonization, their original names from the Freshwater macroinvertebrates checklist and their harmonized names. Reasons of this harmonization are specified (synonymy, wrong spelling, subgenus).



Link: https://doi.org/10.3897/mbmg.9.137772.suppl1

## Supplementary material 2

### Percentage of genetic sequences depending on the type of genes

Authors: Paula Gauvin, David Eme, Isabelle Domaizon, Frédéric Rimet

Data type: pdf

Explanation note: Only the genes representing at least 0.1% of the sequences are shown. All the others genes are grouped in the "Other genes" bar. The numbers above each bar show the total number of sequences retrieved from BOLD affiliated to each gene.



Link: https://doi.org/10.3897/mbmg.9.137772.suppl2

## Supplementary material 3

### Percentage of sequences depending on the countries

Authors: Paula Gauvin, David Eme, Isabelle Domaizon, Frédéric Rimet

Data type: pdf

Explanation note: Only the countries representing at least 0.2% of the sequences are shown. All the others countries are grouped in the "Other countries (133)" bar. The numbers above each bar show the total number of sequences retrieved from BOLD affiliated to each country.



Link: https://doi.org/10.3897/mbmg.9.137772.suppl3

## Supplementary material 4

### Number and names of species associated with BINs

Authors: Paula Gauvin, David Eme, Isabelle Domaizon, Frédéric Rimet
Data type: xlsx
Explanation note: 267 BINS are shared by 410 species.

Link: https://doi.org/10.3897/mbmg.9.137772.suppl4

## Supplementary material 5

### Taxonomic composition of species sharing the same COI haplotype after alignment with fwhF2/fwhR2N primers

Authors: Paula Gauvin, David Eme, Isabelle Domaizon, Frédéric Rimet
Data type: xlsx
Explanation note: A total of 57 haplotypes shared among 116 species were identified.

Link: https://doi.org/10.3897/mbmg.9.137772.suppl5